DOCUMENT RESUME

ED 073 158 TM 002 406

AUTHOR	Lord, Frederic M.
TITLE	Variance-Stabilizing Transformation of the Stepped-Up Reliability Coefficient.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	ETC-RB-72-48
PUB DATE	Oct 72
NOTE	7p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Variance; *Mathematical Models; Research; *Statistical Analysis; Technical Reports; *Test Reliability; *Transformations (Mathematics)

ABSTRACT

The stepped-up reliability coefficient does not have the same standard error as an ordinary correlation coefficient. Fisher's Z -transformation should not be applied to it. Appropriate procedures are suggested. (Author)

ED 073158

TM 002 406

RESEARCH BULLETIN

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RB-72-48

VARIANCE-STABILIZING TRANSFORMATION OF THE STEPPED-UP RELIABILITY COEFFICIENT

Frederic M. Lord

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
October 1972

Variance-Stabilizing Transformation of the Stepped-Up Reliability Coefficient

Frederic M. Lord

Educational Testing Service

Abstract

The stepped-up reliability coefficient does not have the same standard error as an ordinary correlation coefficient. Fisher's z - transformation should not be applied to it. Appropriate procedures are suggested.

Variance-Stabilizing Transformation of the Stepped-Up Reliability Coefficient[1]

Frederic M. Lord

Educational Testing Service

The stepped-up reliability coefficient R considered here is given by the familiar Spearman-Brown formula

$$R = \frac{2r}{1 + r} \tag{1}$$

where r is the observed product-moment correlation between two supposedly parallel sets of measurements X_1 and X_2; or, perhaps better, where r is the maximum likelihood estimate of their correlation under the assumption that X_1 and X_2 are bivariate normal and have equal population variances (Jackson & Ferguson, 1941, eq. 85). Of course, R is the estimated reliability of $X_1 + X_2$. Although R is an estimate of a product-moment correlation coefficient, it is not itself a product-moment correlation and consequently does not have the frequency distribution and the sampling variance of a sample product-moment correlation.

For either definition of r, assuming bivariate normality, as we shall throughout, the large-sample variance of r is

$$\text{Var } r = (1 - \rho^2)^2/N \quad , \tag{2}$$

where ρ is the population correlation. The large-sample variance of R is easily found from (1) and (2) by the "delta" method (Kendall & Stuart, 1958, section 10.6) to be

$$\text{Var } R = 4(1 - P)^2/N \quad , \tag{3}$$

where $P = 2\rho/(1 + \rho)$ is the population value of R. Kristof (1963) has shown the exact sampling variance of R to be

$$\sigma_R^2 = \frac{4(N - 1)(N - 2)}{(N - 3)^2(N - 5)} (1 - P)^2 \quad .$$

Since R is not normally distributed in samples of typical size, research workers sometimes apply Fisher's z-transformation to R and assume that the transformed value has a variance of $1/(N - 3)$ regardless of the value of P. This is incorrect. The large-sample variance of $z_R \equiv \frac{1}{2}[\log(1 + R) - \log(1 - R)]$ is found to be $4/N(1 + P)^2$. This is almost always larger than $1/(N - 3)$. It is not independent of P.

The variance-stabilizing transformation for R can be found from (3) by a standard procedure (Kendall & Stuart, 1958, Exercise 16.18; Eisenhart, 1947):

$$Z = \int^R (N \text{ Var } R)^{-\frac{1}{2}} \, dP$$

$$= \tfrac{1}{2} \int^R \frac{dP}{1 - P}$$

$$= -\tfrac{1}{2} \log(1 - R) \quad . \qquad (4)$$

The large-sample variance of Z is $1/N$, as required, regardless of the value of P. Rewriting Z in terms of r shows, as should be expected, that

$$Z = -\tfrac{1}{2} \log\left(1 - \frac{2r}{1 + r}\right)$$

$$= \tfrac{1}{2}[\log(1 + r) - \log(1 - r)] \quad , \qquad (5)$$

which is simply Fisher's z -transformation for r . Conclusions reached from a study of suitably transformed R must be the same as those from a study of suitably transformed r .

Kristof (1964) has given a large-sample, likelihood ratio test for the case where just two values of R are to be compared. Where two or more values of R are to be compared, they can be transformed by (4) or (5) and each treated (possibly in an analysis of variance) as normal with variance $1/(N - 3)$. This procedure will have good properties in samples of moderate size, at least in the case where r is the sample product-moment correlation, since such properties have been demonstrated for Fisher's z -transformation. This procedure might be applied, for example, to data studied by Traub and Hambleton (1972).

References

Eisenhart, C. Inverse sine transformation of proportions. In C. Eisenhart, M. W. Hastay, & W. A. Wallis (Eds.), Techniques of statistical analysis. New York: McGraw-Hill, 1947. Pp. 397-416.

Jackson, R. W. B., & Ferguson, G. A. Studies on the reliability of tests. Bulletin No. 12. Toronto: University of Toronto, 1941.

Kendall, M. G., & Stuart, A. The advanced theory of statistics. Vol. 1. New York: Hafner, 1958.

Kristof, W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. Psychometrika, 1963, 28, 221-238.

Kristof, W. Testing differences between reliability coefficients. The British Journal of Statistical Psychology, 1964, 17, 105-111.

Traub, R. E., & Hambleton, R. K. The effect of scoring instructions and degree of speededness on the validity and reliability of multiple-choice tests. Educational and Psychological Measurement, 1972, 32, 737-758.

Footnote

[1]Research reported in this paper has been supported by grant GB-32781X from the National Science Foundation.